How Basic is (Patented) University Research? The Case of GM Crops*

Maurizio Conti, Pierre Régibeau and Katharine Rockett

University of Essex and CEPR

June 2003

Preliminary

*The authors would like to acknowledge the generous support of ESRC grant L145251003 for this project.

Abstract

One of the main reasons for subsidising university research is the widespread belief that it generates proportionally more positive knowledge externalities than corporate research. Over the last two decades, however, this belief has been shaken by the increasingly aggressive patenting of university-based innovation. . This perception was supported by Henderson, Jaffe and Trajtenberg (1998) who found both a sharp increase in university patenting and a decrease in the relative 'importance' of university innovation over the later part of their 1965-1992 sample. In this paper, we have compared the knowledge externalities generated by university and corporate patents related to GM crop research. Our main measure of knowledge externalities is the total number of third party cites generated by a patent.
Our main result is that patented university research is not associated with greater knowledge externalities than corresponding corporate patents. If anything, corporate patents appear to generate greater numbers of net citations. This basic conclusion survives when we control for a number of variables that could affect citation counts (e.g. patent examiner effects) and when we break our sample into sub-periods. This does not imply that university patents are similar to corporate patents in every respect. We find two main differences. Firstly, there is some evidence that the shape of the distribution of citations is not identical for the two groups of patents as university patents appear to experience a more sluggish start than their corporate brethren. Secondly, even controlling quite narrowly for areas of specialisation, university patents receive a disproportionate number of cites from other university patents. These two results suggest that there are some fundamental differences in the *types* of knowledge flows generated by university and corporate patents.

Economists have long been convinced that innovation is a crucial factor of economic growth. This belief has only been reinforced by the emergence of new theories where growth is explicitly driven by R&D investments and the resulting flows of knowledge. [1] If growth depends on innovation, and growth is deemed to be desirable, then one should worry about the level of investment into innovative activities. In particular, one would like to know whether private investment decisions are likely to lead to levels and types of innovation that are close to the social optimum.

Not surprisingly, the theoretical literature is inconclusive. To the extent that private inventors fail to capture the whole social surplus created by their efforts, innovation will be underprovided. Moreover, if the ratio of social to private benefits varies across fields of research, the mix of inventions provided by the market will also be sub-optimal. On the other end, private provision of innovation might also be too high. This is essentially a 'excess entry' phenomenon: individual inventors do not consider the fact that their own successful innovation might hurt the profits of their competitors. This negative pecuniary externality – known as the 'business stealing' effect – leads to too much 'duplication; of R&D effort

In spite of the agnosticism of economic theory, policy-makers clearly believe that current levels of innovation are too low. While public initiatives aimed and promoting the 'knowledge-based economy' flourish, it would be hard to find a single policy explicitly aimed at reducing the amount of resources devoted to innovation. University research has long been an important component of innovation policies. This is not only true in the US, where federal research grants account for a substantial proportion of the budget of leading universities but also in several European countries where university funding is increasingly linked to measures of research outputs. State financing of university research is mostly justified by two types of arguments. The first one refers to the complementarity between teaching and research, alleging that quality higher education can only be provided by individuals who are themselves at the forefront of their field. The second argument is that university research is more 'basic' than corporate research and, as such, involves greater benefits for society at large.

[1] For a review of this literature, see Aghion and Howitt (1998).

Trajtenberg, Henderson and Jaffe (1997) have pointed out that this notion of 'basicness' is linked to at least three logically distinct ideas. A first possible interpretation of 'basic' is that universities are, on average, involved in more path-breaking research than the corporate sector.   One can also interpret 'basic' as 'fundamental' or 'important': universities are more likely to create new fields that are subsequently developed by the private sectors. As those fields – and the associated social benefits – would not have emerged (or at least not as fast) without university involvement, the total benefits associated with this 'upstream' research are larger than the benefits associated with the more applied 'downstream' research. Finally the degree of 'basicness' of an innovation might also refer to its degree of appropriability: the more basic the research the lower the proportion of total surplus that can be secured by the inventor.

These three aspects of 'basicness' might be logically distinct but, from a policy point of view, they are inextricably linked. The reason why policy-makers care about 'basic as path-breaking' is because it is associated with 'basic as important'. Moreover, the reason why policy-makers consider subsidising university research is that they believe that these two notions of 'basicness' are closely correlated with 'basicness as lack of appropriability'. It is because university research is thought to be different, more important and less appropriable than corporate research that it receives special treatment from the State. Until the late 1970s this line of argument was generally accepted. As, by law, US universities could essentially not directly benefit from the commercialisation of their research, traditional academic career incentives were bound to bias university innovation towards the more 'basic' kind. In fact, Trajtenberg, Henderson and Jaffe (1997) exploited this *prior belief* that university research was more basic to develop measures of the three aspects of 'basicness' discussed above.

In 1980, however, the Bayh-Dole Act allowed universities to retain property rights on inventions that had benefited from federal funding. The universities ' control over their intellectual property was further strengthened in 1984. These legal changes led to a huge increase in university patenting.[2] As universities started selling the output of their research and emerged as potential rivals to the private sector, the validity of the traditional argument for supporting university research began to be questioned. To address this concern, Henderson, Jaffe and Trajtenberg (1998) examined whether the degree of 'basicness' of (patented) university research actually changed after the legislation of the early 1980s. Using a random sample from all patents assigned to universities between 1965 and 1992, they found that, although the difference has declined over time, university patents were still more general (basic as breakthrough) and more important than the corresponding corporate patents.

We revisit this issue with more recent data on GM crop patents. Bio-technology in general, and GM crop research in particular, is a poster child of the controversy surrounding university patenting. Over the last fifteen years, biotech has seen a high proportion of university patents, university-corporation joint ventures and numerous researcher leaving the academic ranks to joint private companies or found their own. If one is likely to find a strong convergence between university and corporate research universities in any field of research, biotech is a prominent candidate. In other words, limiting ourselves to GM crop innovations provides a strong test of the convergence hypothesis: if it is rejected in this sample, it is quite likely to be rejected in most other research area.

Using patent data for a well-defined research field has other advantages. Firstly it allows for better control of research area than the three-digit patent classes traditionally used in cross-sectional data sets. Secondly, it makes it easier to identify the start of new innovation cycles. As citation patterns might vary along such cycles and universities might be more or less involved in different phases, this is potentially important. GM crop innovations are especially convenient in this respect as we can examine patents from the very beginning of the field. Finally, public policy towards university research can be – and is – tailored to specific research areas. In particular, the government controls the allocation of federal funds and can decide whether or not biotechnology, or even GM crops, is a field of research that it cares to subsidise at the

[2] See Henderson, Jaffe and Trajtenberg (1998)

university level. It is therefore useful to check whether the conclusions drawn from larger data sets are confirmed for more narrowly defined domains of innovation.

Our main conclusion is that, over the period eleven year period from the beginning of 1988 to the end of 1998, university GM crop patents have not proved to be more important than corporate patents. If anything, the evidence suggests that university patents might, on average, have generated fewer 'knowledge externalities' than their corporate equivalent. This basic conclusion does not change when we consider the role of patent examiners and attorneys, when we control for precise areas of research or when we look at various sub-periods within the sample.

Still, university and corporate patents are not identical. We find that the shape of the cite distributions for university and corporate patents might be different. University patents seem to need more time to start gathering citations. As our citation data is censored, this difference in citation profiles would bias the estimate of the importance of university patents downwards. However, correcting for this asymmetry does not suffice to reverse our basic conclusion. University patents also receive cites disproportionately from other universities. This effect remains significant even when we control for the precise area of research for which the patents are obtained, so that it cannot be explained by a common pattern of specialisation among universities. This suggests that the explanation must be related to the flow of information between different types of agents: university researchers are simply more aware of the (patented !) work done by other universities. Finally we find that, although university patents appear to rely more heavily on published scientific research than corporate patents, this effect disappears once one distinguishes between references to third party papers and references to papers (co-authored) by one of the inventors. Hence university patents are complementary products of the inventors' own academic research but they do not draw on the overall pool of 'basic' scientific knowledge more heavily than corporate patents.

The rest of the paper is organised as follows. Section 1 presents the data. The net citation patterns of university and corporate patents are compared in section 2, while section 3 explores further differences between the two types of patents. Section 4 concludes.

1. Data

We use US patent data obtained from the USPTO web site. We focus on patented innovations about genetically modified crops. These are defined as inventions concerning the transformation of crop plants through genetic engineering techniques. Our data set begins with the first GM crop patent, granted in 1988 and stops at the end of 1998. There are a total of 635 patents, 150 of which were obtained by universities. Given our extensive search process[3], we are confident that we have information about nearly all of the relevant patents granted over this period of time. For each patent, we encoded the information routinely available from the USPTO site: date of filing, date of grant, assignee, main patent classes, number of claims, number of US and foreign patents cited, number of scientific articles cited, number of citations obtained from subsequent patents, name of the primary patent examiner and name of the attorney or agent shepherding the patent application through the USPTO approval process. In the current version of the paper, the number of citations received by each patent is counted as of February 2003. This means that, for the last patent in our data set, more than four years have elapsed between their date of grant and the last possible date of citation.

This standard set of information was enlarged in several directions. For each patent, the identity of the citing patents was established. This enabled us to identify own citations, citations coming from distinct fields of invention and citations made by patents held by universities. We also counted the number of scientific papers referenced in each patent, separating references to articles written by one of the inventors from references to articles written by others. Finally, we read the abstract and the claims of each of the patents[4] in order to determine the precise object of the patented invention. We know therefore whether a patent claims the modification of a specific trait (e.g. improved nutrition or herbicide resistance) and/or the modification

---

[3] We started by checking all patent titles within the 800 class. We then ran abstract searches on a large combinations of key words such as ‘gene & plant’, ‘gene & specific plant’, ‘transgenic’ and so on. Each likely candidate was then read to determine whether or not it belonged in the sample. In many cases, this was the only manner of discriminating between genetic transformation through biotechnology and traditional methods of plant husbandry. As a further check we also looked at all of the *citing* patents that we did not already have in our data set.

[4] When necessary, we also referred to the sections ‘background of invention’ and ‘Description of the Innovation’.

of a specific plant. We also know whether a patent claims a process and/or a specific biological construct (e.g. a gene, a DNA sequence).

Our main goal is to compare the knowledge externalities of university and corporate GM crop research. While looking at *patented* innovation is in itself of some interest, one would hope that the answers obtained on such a sample would also tell us something about the overall research of the two types of institutions. This point is illustrated in figure 1. (also found in Trajtenberg at al. (1997)).

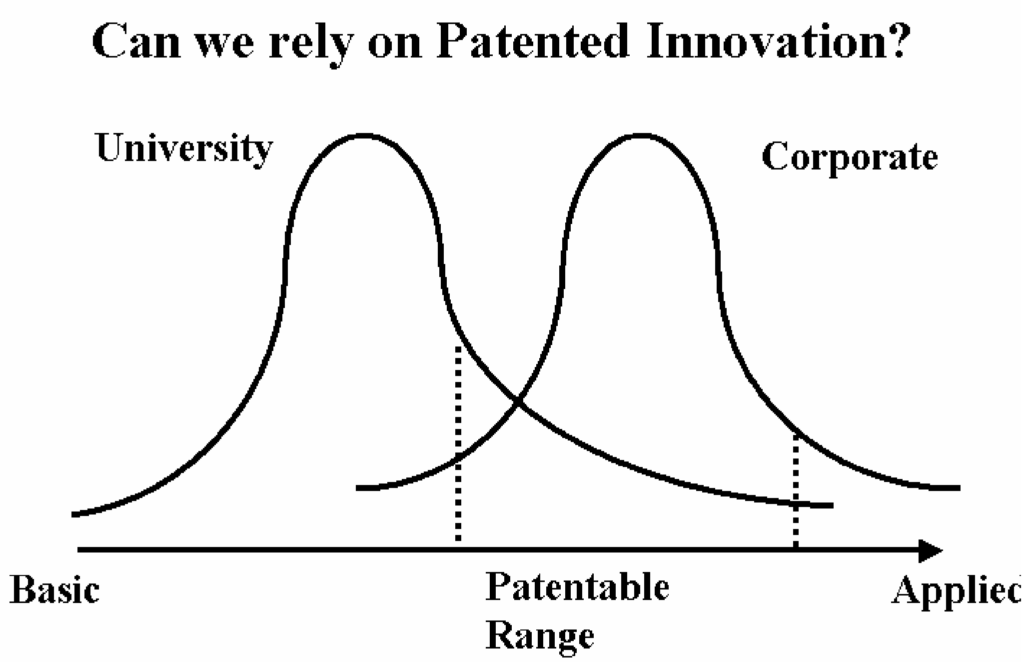

Figure 1

The two curves represent the distribution of university and corporation innovative activities going from more to less basic (in any sense one wishes). As patents are awarded neither to mere concepts or ideas nor to minute changes in the state of the art, there is an intermediate range of innovations that are patentable. If universities do indeed have a distribution of innovations that is weighted more heavily towards ‘basicness’ then this bias ought to show within the subset of *patentable* innovations. If we further assume that there is no university-specific bias in the ypes of innovations that are actually patented, then the ‘basicness’ bias should also appear within the subset of *patented* research.

2. Knowledge Externalities

In this section we ask whether there is any evidence that GM crop innovations patented by universities generate more external benefits than innovations generated by other types of agents. Externalities will be larger the more important the innovation and the lower the degree of appropriation of the innovation 's benefits by the patent-holder. A widely used measure of 'importance' is the number of downstream cites that a patent generates. Since later patent applications are, in theory at least, required to list all relevant sources of prior art, one would expect that patents that help develop significant sub-fields of research will, other things equal, be more widely cited. Following Trajtenberg, Henderson and Jaffe (1997) we also consider self-citations, i.e. cites found in subsequent patents obtained by the same agent, as an indication that the innovator is internalising the benefits of his initial innovation. Our measure of external benefits will therefore be the total number of *net* cites, obtained by subtracting own citations from the total number of cites[5]. This measure is represented by the variable CITE.

We are interested in the following relationship:

$$CITE = \alpha + \beta X + \gamma TYPE$$

, where X is a vector of control variables and TYPE is a vector of dummy variables capturing the identity of the agent to whom the patent is assigned. We will include four such dummies. UNIV takes the value 1 if the assignee is a university and 0 otherwise. Similarly, GOV identifies a government agency and INDIV indicates that the patent-holder is an individual. We also use a MIXED variable that corresponds to private corporations that are located in the same town as a university that has a major agro-biotech research programme. By default, then patent holders that are not identified by any of these four dummy variables are private corporations.

Given the nature of our data, our dependent variable will be heavily censored. This is especially true in this industry where citations appear to be less skewed

[5] |Contrary to Trajtenberg et al. (1997) we only include first generation cites. Over a period of just 11 years this is unlikely to make much of a difference. In fact, Trajtenberg et al. themselves find that varying the weight placed on second period citations does not significantly affect their results.

towards the early years than in other samples.[6] To understand the nature of this censoring, we must refer to basic principles of Patent Law. Over the period covered by our data, the US still has a 'first to invent' system without any disclosure of patent applications. As patent applications are kept secret until the patent is actually obtained, the date at which the patent is granted determines when it first becomes 'available' for citation. Clearly then, the later a patent is granted, the less likely it is that it will have gathered a large number of cites. The number of cites will also depend on the status of the patent in terms of 'prior art'. In a first to invent system, priority is determined by the date of invention. Although an inventor can, in principle, wait for a significant period of time before actually filing a patent applications, there are reasons to believe that, in practice, the filing date should be a reasonable approximation of the date a which the invention was actually reduced to practice. In particular, the Law includes a number of 'statutory bars' that would make it rather difficult to wait a long time before invention and filing.[7] For a given grant date, a later filing date means that the patent will become legal precedent for fewer of the applications that are still in the process of being approved. In other words, the later the file date, the lower the number of expected cites.

To account for the censoring we will therefore include both the filing date and the grant date of a patent as control variables. Defining FILE and GRANT as the filing and grant dates measured in months from a common origin, we run a negative Binomial regression between CITE and these two dates:

$$\underset{}{\text{CITE}} = \underset{(0.613)}{6.784^{***}} + \underset{(0.0069)}{0.00105}\ \text{FILE} - \underset{(0.008)}{0.0371^{***}}\ \text{GRANT}$$

,where the marginal effect of FILE is 0.00359 and the marginal effect of GRANT is –0.1. The coefficient of GRANT is significantly negative, as expected. It is also not small as a patent granted one year later will on average get one fewer cite. In a sample where the median number of cites is 3, this is not negligible. The coefficient of FILE is not significant. This, actually, is not surprising because we would expect FILE to pick up two effects of opposing sign. On the one hand, as explained above,

[6] See, for example Caballero and Jaffe (1993)

[7] For example, "If an inventor is the first to invent but waits more than a year to apply for a patent during which time someone else makes the same discovery and describes, publicly uses, sells or patents

an earlier filing date means higher ‘priority’ status and should therefore translate into more cites. On the other hand, more important innovations tend to arise relatively early in a given innovation wave. This would show as a positive correlation between CITE and FILE.

We can now have our first look at university patents. To do this we keep FILE and GRANT as control variables and introduce the full set of assignee dummies described above. As CITE is a non-negative count variable and is often small. Under such conditions, OLS is unlikely to perform well. In Table 1, we compare OLS to the results obtained by using both Poisson and negative binomial regressions. As our data clearly suffers from heteroskedasticity, the numbers presented in brackets are robust standard errors. The estimates for the Poisson and negative binomial regressions are regression coefficients, not marginal effects. Their magnitude is therefore not directly comparable to the size of the OLS estimates.

Table 1

Dep: CITE	OLS (Robust)	Poisson (Robust)	Negative Binomial (Robust)
Grant	-0.158*** (0.329)	-0.0211*** (0.0575)	-0.0302*** (0.00427)
File	-0.71^{+} (0.45)	-0.00611 (0.00589)	-0.00464 (0.00391)
University	0.394 (0.96)	0.0491 (0.176)	-0.179^{+} (0.126)
Mixed	0.443 (0.889)	0.0645 (0.143)	-0.252** (0.116)
Government	-1.181^{+} (0.735)	-0.55** (0.260)	-0.596*** (0.2119)
Individuals	-1.959 (1.86)	-0.451 (0.428)	-0.506* (0.291)
Constant	38.03*** (4.275)	5.335*** (0.309)	6.6*** (0.319)
	R^2 = 0.314	Pseudo R^2 = 0.344	Prob > X^2 = 0.000
		g.o.f: p = 0.000	$^{+}$ = nearly significant at the 0.1 level

These results certainly do not support the idea that university patents generate more net cites than corporate patents. In fact, the coefficient of the university dummy obtained in the negative binomial regression is negative and almost significant at the 0.1 level. Interestingly, once appropriate techniques for count data are used, the coefficient of the government dummy is quite significantly negative. Table 1 also

the invention in this country, the invention is unpatentable by the first inventor, even though he is truly demonstrates that accounting for the discreteness of the dependent variable matters. Moreover, as our data show strong signs of over-dispersion, the negative binomial approach appears to be more appropriate. We will therefore limit ourselves to reporting the negative binomial estimate for the rest of the paper.[8]

In Table 1, we controlled for the dates of filing and grant by using linear variables corresponding to the number of months elapsed since a fixed common origin. To allow for non-linearities in the patent citation profile, we also estimated equations when FILE and GRANT are replaced by year dummies. The results are shown in Table 2. Although the coefficients of the year dummies confirm that the relationship between CITE and the dates of filing and grant is not perfectly linear, the coefficients of the assignee dummies remain pretty much the same. Government agencies, individuals and firms with close links to universities obtain patents that generate significantly fewer net cites than corporate patents. The only novelty is that, when year dummies are used to control for both grant and file dates, the coefficient of the university dummy is now significantly negative at the 0.1 probability level.

Table 2

Dep: CITE	**Negative Binomial**	**Negative Binomial**	**Negative Binomial**	**Negative Binomial**
Grant	-0.0302*** (0.00427)	-0.029*** (0.00395)		
File	-0.00464 (0.00391)		-0.00847** (0.00374)	
University	-0.179^{+} (0.126)	-0.162 (0.123	-0.168 (0.1222)	-0.196* (0.1203)
Mixed	-0.252** (0.116)	-0.251** (0.1107)	-0.203* (0.1096)	-0.223** (0.1084)
Government	-0.596*** (0.2119)	-0.635*** (0.2096)	-0.591*** (0.2249)	-0.635*** (0.2247)
Individuals	-0.596* (0.291)	-0.564** (0.2696)	-0.483* (0.2859)	-0.543** (0.2681)
File Dummies		Yes		Yes
Grant Dummies			Yes	Yes
Constant	6.6*** (0.319)	4.80*** (0.773)	1.56*** (0.525)	-0.355 (0.367)
Prob > X^2	0.000	0.000	0.000	0.000

the one who invented first", Miller and Davis (1983).

[8] Every reported regression has also been estimated using Poisson. The results are never markedly different.

So far, we have obtained fairly strong evidence that University GM crop patents are *not* associated with larger externalities, as measured by net downstream cites. In the rest of this section, we will test the robustness of this conclusion by controlling for factors that might conceivably bias our estimates.

One possible source of bias comes from the nature of the patenting process. Effectively, the US PTO 's "policy" is just the result of thousands of bilateral interactions between applicants and their representative and patent examiners. Applicants or, more likely, their legal representatives control the content of the initial submission and decide how to respond to the examiner 's requests. Examiners have a fair amount of discretion as to how the standards of novelty, non-obviousness and usefulness apply to specific claims. Clearly then, the two parties have some influence on the precise content of the final patent. This influence might affect the number of claims that are finally accepted, the generality of these claims and the extent to which the patent acknowledges 'prior' art. The first two factors can significantly affect the legal protection granted to a given invention and can, therefore, also influence the flow of citations received by the patent. If each examiner or each firm/legal agent only handles a very small proportion of the patents in the data set, we would not expect their idiosyncrasies to bias the coefficient of the university dummy. After all there is no reason to believe that universities would systematically choose less aggressive IPR lawyers to represent them or that they would systematically be paired with tougher PTO examiners. In our sample however, several Law firms and, especially, a few examiners handle a large share of patents. Just nine law firms account for more than 40% of the patents, while the top five examiners account for 69% of the cases. Since the total number of patents involved is also relatively small, one cannot a priori rule out some kind of examiner or attorney-related bias. We have therefore estimated equations that include examiners and/or attorney dummies on the right-hand side. We assigned a dummy to each agent involved in at least 10 cases.

As can be seen in Table 3, some of the examiners and attorney dummies are highly significant. This supports the hypotheses that the 'give and take' involved in the patent approval process leaves room individual influences[9]. On the other hand,

[9] Of course, examiners and attorneys might be narrowly specialised the GM crop area. In that case, the significant coefficients might simply reflect the fact that examiner 8, for example, deals with types of

the coefficients of the assignee dummies do not change much. In particular, the university coefficient is still negative and, in some case, significantly so.

Table 3

	Net Cites	Net Cites	Net Cites
Constant	6.509***	6.735***	6.614***
Grant	-0.0309***	-0.0311***	-0.0317***
File	-0.0049	-0.0043	-0.00485
University	-0.127	-0.222*	-0.191^{+}
Mixed	-0.187^{+}	-0.535***	-0.503***
Government	-0.513**	-0.652***	-0.576***
Individuals	-0.446^{+}	-0.561*	-0.531*
Exa1	0.202		0.311*
Exa2	-0.039		-0.013
Exa3	0.0204		0.051
Exa4	0.553***		0.577***
Exa5	-0.0316		-0.13
Exa6	-0.432^{+}		-0.341
Exa7	0.36		0.242
Exa8	0.56***		0.621***
Exa9	-0.0093		0.019
Exa10	0.0792		0.081
Atto1		-0.54***	-0.54***
Atto2		0.0144	0.087
Atto3		0.394	0.605*
Atto4		0.127	0.03
Atto5		0.725**	0.72**
Atto6		0.416**	0.51***
Atto7		-0.101	0.043
Atto8		0.336	0.539
Atto9		0.191	0.333
Atto10		-1.29***	-1.30***
Atto11		-0.349	-0.191
Atto12		-0.397	-0.282
Atto13		0.134	0.164
Atto14		-0.24	-0.182
Atto15		-0.616***	-0.564**
Atto16		0.218	-0.276
Atto17		-0.048	-0.013
Atto18		0.0264	-0.077
Atto19		0.273	0.345
Prob > X^2	0.000	0.000	0.000

Although university patents tend to generate fewer ‘externalities’ over the whole sample, our results might hide some evolution of the universities ‘ behaviour as the new ‘GM’ technological wave evolves. One might for example believe that universities disproportionately contribute to the initial ‘basic’ innovations that launch

innovations that are more important than, say, examiner 6. This, however, is not the case. Examiner and attorney dummies remain significant even if one controls quite precisely for the type of innovation a new field but that, as this field evolves, universities turn their attention elsewhere, merely taking patents corresponding to fairly routine graduate student ‘ s research.[10] To evaluate this hypothesis, we re-estimated the same relationship as in Table 1, breaking the sample into an ‘early’ and a ‘late’ period, based on the filing date. We used a variety of dividing dates between January 1992 and January 1996. In all of these regressions the coefficient of the university dummy was not significantly different from zero. The only cases where the coefficient was nearly significant at the 0.1 level were cases where the coefficient was negative. We must therefore conclude that there has not been any period over which university patents were associated with greater cite externalities than corporate patents.

So far, we have implicitly assumed that the distribution of citations over time has a similar shape (but not necessarily a similar ‘height’) for all types of assignees. However, if university patents were to embody more drastic innovations than corporate patents they might initially receive citations more slowly as other inventors might need more time to follow in the university ‘s footsteps. Such a situation is shown in figure 2 which represents the distribution of cites overtime for two patents granted at the same time. Although the total lifetime total of cites would be the same for both patents, the distribution of cites for the university patent peaks later. Given that our citation data is censored at t_s, however, we would erroneously conclude that university patents give rise to fewer net cites.

In order to test for such a difference between the distributions of cites we would need information about the timing of all cites for all patents, something that we do not currently have. What we have, for each patent, is the date of grant of the first citing patent. We can therefore compute the citation lag (CILAG) as the difference between that date and the date at which the cited patent itself was granted. The citation lag can then be treated as a ‘survival’ time with the first citation of a patent spelling the ‘death’ of this observation. The determinants of such lags are commonly estimated by using the censored Weibull regression.

---

involved.

[10] For example, isolating a useful gene was once frontier research but it has now become a typical subject for an average doctoral Thesis.

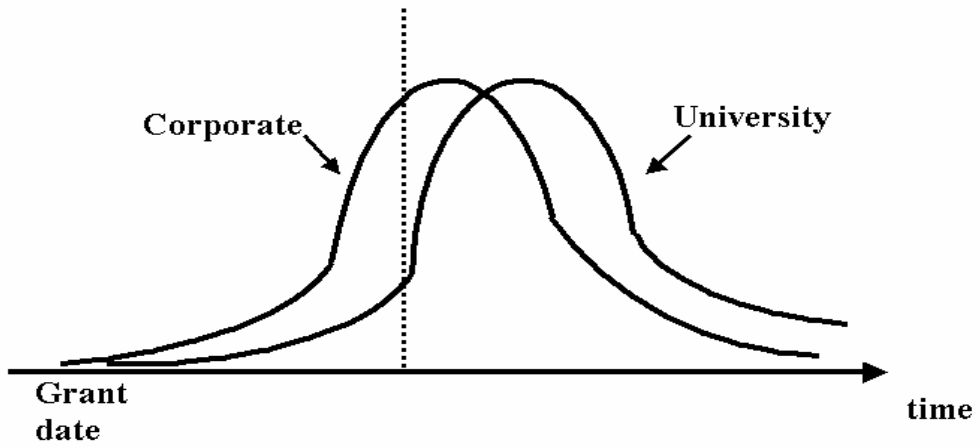

Figure 2

Right-hand side variables include the assignee dummies and two control variables. The date of grant of the cited variable corrects for the fact that our 'survival' experiments begin at different dates depending on when the cited patent is granted. We also try to control for differences in the lifetime citations that would be received by the patents. If we do not, then a patent with more lifetime citations but the same shape of citation distribution would tend to have a shorter lag to its first received citations. As a rough control, we run the regressions with the total number of net cites on the right-hand side. As explained above, we would expect larger values of he cite variable to be associated with shorter citation lags.

Table 4 presents the results for both the censored Weibull and the censored normal regression. The coefficients shown for the Weibull regression are time ratios. As expected, later grant dates and higher total number of cites are associated with smaller citation lags. The university dummy has a significantly positive coefficient in both regressions, lending support to the hypothesis that university patents might have pattern of citation that begins more slowly. The estimated coefficient in the censored normal regression suggests that the difference between the distributions amounts to a little more than 4 months. According to the coefficient of the Weibull regression the citation lag for university patents is, on average, 9.4% longer than for corporate patents. As the average citation lag is about 4 years, this amounts to an additional delay of about 5 months.

Table 4

Dep: CILAG	**Censored Normal**	**Censored Weibull (Time Ratio)**
Constant	82.089*** (5.135)	
Grant	-0.276*** (0.0319)	0.992*** (0.00084)
Cite	-0.595*** (0.0728)	0.984*** (0.0156)
University	4.148** (1.677)	1.094** (0.0446)
Mixed	1.744 (1.835)	1.0267 (0.0456)
Government	9.26** (3.86)	1.21** (0.113)
Individual	-7.29* (4.265)	0.8394* (0.0863)
	0.000	0.000

Using CITE to control for the number of lifetime cites is clearly not completely satisfactory since whether CITE is a good measure of lifetime cites depends itself on the shape of the distributions and therefore on the citation lag. On broad cross-sectional patent data set, the number of claims in a patent has been shown to be a useful alternative measure of the importance of a patent.[11] As shown in the third column of table 4, replacing CITE with the total number of claims does not change the size or significance of the university dummy coefficient much.

To evaluate the possible effect of these differences in the pattern of citation we modified our data by adding five months to the grant date of university patents. In other words, since university patents appear to suffer an additional lag of five months, we treat them as if they had in fact been awarded five months later than they actually were. We then re-estimated the relationship between cites, file and grant dates, assignee dummies, with and without examiner and attorney dummies. The results of these negative binomial regressions are presented in table 5. Robust standard errors are shown in brackets).

Comparing table 5 to tables 1 and 3, we see that using the 'handicapped' grant dates does, as expected, increase the coefficient of the university dummy. Still this coefficient is still negative (but not significantly different from zero). The possible asymmetry in the shape of the distributions of cites is not enough to reverse our basic conclusion.

Table 5

	CITE	CITE
Constant	**6.6*** (0.318)**	**6.61*** (0.379)**
File	**-0.0046 (0.00392)**	**-0.0049 (0.0036)**
Grantcorr	**-0.0302*** (0.00427)**	**-0.0317*** (0.0418)**
University	**-0.0274 (0.129)**	**-0.0328 (0.13)**
Mixed	**-0.252** (0.116)**	**-0.503*** (0.159)**
Government	**-0.596*** (0.211)**	**-0.576*** (0.212)**
Individuals	**-0.506* (0.291)**	**-0.531* (0.292)**
Examiner Dummies		**YES**
Attorney Dummies		**YES**
Pseudo R^2	**0.091**	**0.107**

3. Are University Patents Different from Corporate Patents?

So far, our results support the claim that, in the GM crop field at least, university patented innovation does not generate greater knowledge externalities than corporate patents. This seems to suggest that public funding of this type of university research might not be justified. However, such a conclusion might not be warranted. In this section we try to determine whether University patents differ in other aspects and, if so, whether these difference might justify the subsidisation of University research.

[11] See Lanjouw and Schankerman (1999)

The first hypothesis that we explore is that universities might have specialised in different types of innovations within the GM crop field. One possibly important distinction is between patents that claim new *methods* and patents that claim a new gene[12]. To account for this difference, we construct two variables, METHOD and GENE. METHOD is equal to 1 if the patent claims a new method or process and is equal to 0 otherwise. Similarly, GENE is set equal to 1 when new biological entities are claimed and is equal to zero otherwise. Since patents can claim several aspects of the same innovation, these two categories are not exclusive.

In our data set, University patents claim 'Methods' 55 % of the time and 'genes' 61% of the time. The corresponding percentages for corporate patents are 42% and 43% respectively. To the extend that 'method' and 'gene' claims tend to generate different number of net cites than other types of innovations, this might help account for the lesser cite counts of university patents. If, in addition, there were reasons to believe that this type of innovation would be less readily supplied by the market in spite of their significant social benefits, then one would have a plausible case in favour of public funding of university research. To evaluate this argument, we use the two variables as control on the right-hand side of our citation equation. The results of the negative binomial regressions are reported in Table 5. Robust standard errors are in brackets.

Both control variables are significant but have opposite signs. These are rather intuitive. 'Methods' would tend to have broader applicability than most specific genes or DNA sequences. One would therefore expect the number of cites to be larger for 'methods' than for 'genes'. On the other hand, the coefficients of the assignee dummies are barely affected by the inclusion of the two new control variables so that the argument presented above is not supported.

Table 6

	CITE	CITE	CITE
Constant	6.44***	6.68***	6.55***

[12] This category includes all patents that, in their abstract, claim a gene or another piece of biological material such as DNA segments or nucleotides.

	(0.3105)	(0.311)	(0.302)
Method	0.25** (0.1054)		0.193* (0.105)
Gene		-0.353*** (0.0922)	-0.322*** (0.0906)
Grant	-0.0299*** (0.00443)	-0.0294*** (0.00436)	-0.293*** (0.00447)
File	-0.0051 (0.0041)	-0.0045 (0.004)	-0.005 (0.0041)
University	-0.182 (0.128)	-0.182 (0.132)	-0.182 (0.132)
Mixed	-0.290** (0.115)	-0.328*** (0.110)	-0.35*** (0.1102)
Government	-0.621*** (0.209)	-0.658*** (0.199)	-0.67*** (0.199)
Individuals	-0.413 (0.308)	-0.583** (0.311)	-0.506* (0.296)
Prob > X^2	0.000	0.000	0.000

University and corporate research might also differ in the precise areas that they target. Suppose that the proportion of private to social benefits varies across these areas. Private corporations would tend to be more active in the areas with high private to social benefits ratios. Absent any public intervention, then, the mix of research project would not be socially optimal. For example, Harhoff, Régibeau and Rockett (2001) have argued that the faster development of herbicide resistance might be explained by the complementarity between the crop and herbicide interests of some large corporations. Under such circumstances, targeting university research towards the neglected fields would therefore be a sensible policy. This policy would have two interesting implications for the number of citations received by university patents. Firstly, to the extent that private GM crop research is more abundant than university research, this pattern of specialisation would result in larger number of cites for corporate patents than for university patents. This means that, in order to fairly compare the level of citation externalities associated with corporate and university patents, one should control for the precise area of research. If the argument presented has any weight then controlling for fields should increase the value of the coefficient of the university dummy. Secondly, if the argument is correct, the *origin* of citations received should differ significantly, with university patents disproportionately cited by other universities. These two predictions are tested below.

Table 7

	CITE	CITE	CITE	CITE
Constant	6.78*** (0.311)	6.63*** (0.308)	6.85*** (0.306)	6.72*** (0.301)
File	-0.0024 (0.0039)	-0.0029 (0.0041)	-0.0023 (0.0040)	-0.0027 (0.0041)
Grant	-0.032*** (0.00431)	-0.032*** ((0.00445)	-0.0314*** (0.00437)	-0.0312*** (0.0045)
Method		0.238** (0.098)		0.191* (0.098)
Gene			-0.317*** (0.091)	-0.288*** (0.0891)
University	-0.145 (0.119)	-0.145 (0.119)	-0.147 (0.123)	-0.144 (0.122)
Mixed	-0.272** (0.112)	-0.305*** (0.111)	-0.333*** (0.108)	-0.353*** (0.107)
Government	-0.585** (0.231)	-0.60*** (0.232)	-0.633*** (0.22)	-0.640*** (0.223)
Individuals	-0.605** (0.283)	-0.511* (0.300)	-0.666** (0.275)	-0.586** (0.289)
Herbicide	-0.22 (0.167)	-0.178 (0.176)	-0.15 (0.172)	-0.121 (0.178)
Pest	-0.311* (0.165)	-0.30* (0.161)	-0.293* (0.163)	-0.285* (0.160)
Patho	-0.621*** (0.141)	-0.607*** (0.141)	-0.58*** (0.139)	-0.573*** (0.140)
Sugar	-0.38** (0.191)	-0.395** (0.184)	0.324* (0.198)	-0.341* (0.190)
Starch	0.405 (0.268)	0.448 (0.300)	0.478* (0.289)	0.51 (0.315)
Fertility	-0.657** (0.281)	-0.654** (0.270)	-0.60** (0.266)	-0.602** (0.258)
Nutrition	0.25 (0.184)	0.234 (0.183)	0.28(0.190)	0.267 (0.188)
Growth	-1.09*** (0.221)	-1.076*** (0.223)	-1.036*** (0.215)	-1.03*** (0.217)
Prob > X^2	0.000	0.000	0.000	0.000

In tables 7 through 9, we control for the precise field of research by including dummy variables that represent both the specific trait modified and the main plant for which the innovation is claimed. The information required for creating these dummies was obtained by reading the claims of every patent in the data set and, if further clarification was necessary, by consulting the patent sections on the 'technical background and 'description' of the innovation. It is important to realise that not all patents claim the modification of a specific trait and/or of a specific type of plant. In fact patent lawyers often like to state that, when it comes to claims, 'more is less'. By this they mean that the more specific the claim, the narrower its legal scope. This principle explains why most of the coefficients on patent or trait dummies have negative coefficients. Most of the patents for which the dummies take values of zero are patents that do not make trait or plant-specific claims and tend, therefore, to be broader.[13] Still, comparing the coefficients of trait and patent dummies pair-wise, one can readily see that some significant differences emerge (for example between 'growth' and 'Pathogen' or between 'fertility' and 'nutrition'). It does therefore appear that citation rates do vary across sub-fields of research.

On the other hand, controlling for areas of specialisation does not help bolster the coefficient of the university dummy. Comparing with the first column of table 2 and with table 6, reveals a decrease in both the absolute value and the significance of the negative coefficient, but it is very slight.

Table 8

	CITE	CITE	CITE	CITE
Constant	6.695*** (0.3136)	6.544*** (0.3066)	6,794*** (0.303)	6.674*** (0.296)
File	-0.00379 (0.00385)	-0.00421 (0.00404)	-0.0037 (0.004)	-0.0041 (0.0042)
Grant	-0.032*** (0.00417)	-0.0317*** (0.00434)	-0.0312*** (0.0043)	-0.0311*** (0.00444)
Method		0.279*** (0.1005)		0.211** (0.101)
Gene			-0.381*** (0.091)	-0.345*** (0.090)
University	-0.133 (0.121)	-0.151 (0.118)	-0.146 (0.125)	-0.156 (0.122)
Mixed	-0.187 (0.124)	-0.236* (0.123)	-0.286** (0.117)	-0.313*** (0.116)
Government	-0.648*** (0.232)	-0.698*** (0.236)	-0.713*** (0.226)	-0.744*** (0.229)
Individuals	-0.497* (0.297)	-0.391 (0.318)	-0.593** (0.284)	-0.507* (0.30)
Corn	0.357** (0.1504)	0.341** (0.152)	0.293* (0.150)	0.287* (0.152)
Rice	0.834* (0.489)	0.931* (0.523)	0.868 (0.544)	0.943* (0.566)
Soy	0.80*** (0.180)	0.735*** (0.169)	0.914*** (0.175)	0.854*** (0.168)
Wheat	-0.699** (0.322)	-0.791** (0.330)	-0.79*** (0.308)	-0.851*** (0.316)
Cotton	-0.052 (0.211)	-0.0458 (0.232)	0.00378 (0.202)	-0.0004 (0.22)
Potato	-0.115 (0.165)	-0.178 (0.169)	-0.103 (0.167)	-0.151 (0.17)
Tomato	-0.165 (0.170)	-0.226 (0.166)	-0.171 (0.154)	-0.215 (0.153)
Tobacco	-0.320* (0.189)	-0.318* (0.184)	-0.306 (0.190)	-0.302 (0.188)
Pseudo R^2	0.099	0.101	0.104	0.105

[13] Dummies are assigned to the traits and plants that are encountered most often. Hence a number of patents which claim to modify specific, but less frequent, traits or plants are also assigned values of zero for all dummies.

Table 9

	CITE	CITE
Constant	6.80*** (0.3086)	6.74*** (0.2989)
File	-0.0011 (0.0038)	-0.0015 (0.0042)
Grant	-0.34*** (0.00416)	-0.033*** (0.00453)
Method		0.20** (0.0947)
Gene		-0.326*** (0.091)
University	-0.085 (0.117)	-0.114 (0.116)
Mixed	-0.185 (0.115)	-0.283*** (0.1095)
Government	-0.603*** (0.225)	-0.693*** (0.214)
Individuals	-0.55* (0.287)	-0.555* (0.292)
Trait Dummies	Yes	Yes
Plant Dummies	Yes	Yes

The second implication of the 'field of specialisation' story is that universities should tend to operate in a 'world of their own' and, therefore cite each other 's patents disproportionately. To test this we regress the proportion of cites received that originates from universities. Both the numerator and the denominator of the variable UNIPRO are cites received from third parties.[14] For the results presented in Table 10, patents that have not yet received any citations were assigned a value of 0 for UNIPRO. Given the resulting mass point at zero, we used a tobit regression. We also ran the same regressions restricting the data set to patents that have received at least one citation from a third party. The results were not significantly different.

Table 10
Tobit

	Unipro	Unipro	Unipro
Constant	-0.101** (0.0413)	0.702*** (0.183)	0.553*** (0.208)
File		-0.0004 (0.002)	-0.0019 (0.002)
Grant		-0.005** (0.0023)	-0.0041* (0.0024)

[14] Using total cites in the numerator would introduce a spurious positive correlation between this proportion and the university dummy.

University	0.111* (0.0666)	0.115* (0.674)	0.143** (0.074)
Mixed	0.015 (0.0741)	-0.0028 (0.075)	-0.0211 (0.0947)
Government	0.382*** (0.145)	0.445*** (0.146)	0.419*** (0.147)
Individuals	-0.14 (0.163)	-0.198 (0.168)	-0.212 (0.169)
Exa dummies			Yes
Atto dummies			Yes
Prob > X^2	0.000	0.000	0.000

The main conclusion from the first two columns of Table 10 is that universities do indeed receive proportionally more cites from other universities than from the corporate sector compared to what would be expected if citations were allocated randomly across these two types of assignees. Interestingly government patents receive an even greater share of their citations from universities than universities themselves.

However, the pattern of citations identified in the first two columns of Table 10 only establishes that universities have a greater propensity to cite each other. It does not by itself explain where this incestuous pattern of citations comes from. In particular, it does not necessarily imply that it is the result of a common pattern of specialisation. To investigate the issue further, we included our traits and plants dummies as control variables. If the university 's tendency to cite each other can be mostly explain by the fact that they are disproportionately present in the same research areas then controlling for these sub-fields should significantly reduce the positive coefficient of the university dummy. Comparing the last two columns of Table 10 shows that this is not the case. Indeed, the coefficient of the university dummy is both larger and more significant once areas of research are controlled for.

Combining the evidence from tables 9 and 10 strongly suggests that the fact that university patents receive fewer cites than corporate patents cannot be explained by different patterns of specialisation. Accordingly, it would be hard to argue that university research helps correct possible biases in the private sector 's choice of GM crop applications. This also means that the positive coefficient of the university dummy in the UNIPRO regression must reflect some other mechanism. One such mechanism could be the existence of 'social networks' guiding the flow of knowledge, i.e. university inventors might just be more aware of work (included patentable inventions) done by other university scientists.

The results obtained so far are at odds with conventional ideas about university research. Even if one accepts that university patents might not on average look very different from corporate ones, it is hard to believe that the research processes giving rise to these patents are the same. In particular one might think that university patents rely more directly on new advances in the underlying science. To evaluate this idea, we follow Trajtenberg, Henderson and Jaffe (1997) and create a variable that captures the inventors ' propensity to cite papers published in scientific. This variable, ARTPRO, is the ratio between the number of references to published scientific papers and the sum of these references and the number of patents cited. We then regress ARTPRO on the assignee type dummies as well as on the grant and file dates. These two dates are included to control for the fact that the pools of citable articles and patents might have different relative sizes at different stages of the innovation cycle. For example, relevant patents to cite might be in relatively short supply early on. To the extent that the proportion of university to corporate patents vary over time, failure to account for possible time trends in ARTPRO could bias our estimates of the coefficient of the university dummy. As there is a mass point of observation for which ARTPRO is equal to zero, we use a tobit regression. We also restrict the data set to patents that cite at least one reference, be it patent or article. The results are presented in the first column of Table 11. We see that, among all other types of assignees, universities are the only ones who rely significantly more on scientific articles than private corporations.

We also created a measure of own scientific citations defined as the total number of cited papers written by (at least) one of the listed inventors divided by the total number of scientific papers cited. The numerator of this variable was obtained by checking the names of the authors of the scientific articles cited against the names of the inventors of the patented innovation. Given that scientific references do not always include the full list of authors, our variable must be considered as a rather noisy measure of the true ratio.

Table 11 (Tobit)

	Artpro	Own Articles	Third party articles
Constant	4.806* (2.751)	-0.147 (0.106)	0.772*** (0.070)
File	-0.0276 (0.0299)	0.0004 (0.0011)	-0.0012 (0.00077)
Grant	0.0173 (0.0339)	0.00032 (0.0013)	0.00047 (0.00087)
University	2.325** (0.984)	0.180*** (0.0368)	-.0309 (0.0254)
Mixed	0.992 (1.069)	-0.018 (0.0425)	0.0234 (0.0277)
Government	-0.216 (2.334)	0.124 (0.0824)	0.0557 (0.0574)
Individuals	-3.08 (2.305)	0.197** (0.0825)	0.197*** (0.0574)
Prob > X^2	0.000	0.000	0.0031

By regressing OWNARTPRO against the assignee type dummies, we can get an idea of whether university patents rely more on the 'basic' scientific knowledge of the inventor than patents obtained by other assignees. As shown in the second column of table 1, the university dummy is significantly positive. Moreover, if we regress the ratio of articles from third parties cited to the total number of citations (patents and articles) on assignee dummies and time variables, the coefficient of the university dummy becomes negative, albeit not significantly different from zero (see the last column of table 11). Taken together, the results in table 11 suggest that, although university inventors are themselves 'closer' to the basic science underlying their patented innovations they do not rely on the 'basic' scientific knowledge of others any more than corporate inventors.

4. Conclusion

One of the main reasons for subsidising university research is the widespread belief that it generates proportionally more positive knowledge externalities than corporate research. Over the last two decades, however, this belief has been shaken by the increasingly aggressive patenting of university-based innovation. This perception was supported by Henderson, Jaffe and Trajtenberg (1998) who found both a sharp increase in university patenting and a decrease in the relative 'importance' of university innovation over the later part of their 1965-1992 sample. In this paper, we have compared the knowledge externalities generated by university and corporate patents related to GM crop research. Our main measure of knowledge externalities is the total number of third party cites generated by a patent.

Using patent data for a well-defined research field has several advantages. Firstly, it allows for better control of research area than the three-digit patent classes traditionally used in cross-sectional data sets. Secondly, it makes it easier to identify the start of new innovation cycles. As citation patterns might vary along such cycles and universities might be more or less involved in different phases, this is potentially important. Thirdly, public policy towards university research can be – and is – tailored to specific research areas. As the government controls the allocation of federal funds it can decide whether or not biotechnology, or even GM crops, is a field of research that it cares to subsidise at the university level. It is therefore useful to check whether the conclusions drawn from larger data sets are confirmed for more narrowly defined domains of innovation.

Our main result is that patented university research is not associated with greater knowledge externalities than corresponding corporate patents. If anything, corporate patents appear to generate greater numbers of net citations. This basic conclusion survives when we control for a number of variables that could affect citation counts (e.g. patent examiner effects) and when we break our sample into sub-periods. This does not imply that university patents are similar to corporate patents in every respect. We find two main differences. Firstly, there is some evidence that the shape of the distribution of citations is not identical for the two groups of patents as university patents appear to experience a more sluggish start than their corporate brethren. Secondly, even controlling quite narrowly for areas of specialisation, university patents receive a disproportionate number of cites from other university patents. These two results suggest that there are some fundamental differences in the *types* of knowledge flows generated by university and corporate patents. Understanding these differences would require not only data on the complete distribution of citations, which we do not currently have, and some modelling of the knowledge flows. This is something that we intend to turn too very soon.

References

Aghion, P. and P. Howitt, 1998, "*Endogenous Growth Theory*", MIT Press.

Caballero and A.B. Jaffe, 1993, "How High are the Giants ' Shoulders: An Empirical Assessment of Knowledge Spillovers and Creative destruction in a Model of Economic Growth", in O. Blanchard and S. Fischer, eds, *NBER Macroeconomics Annual*, vol. 8., MIT Press.

Jaffe, A.B. and M. Trajtenberg, 1996, "Flows of Knowledge from Universities and Federal Laboratories: Modeling the Flow of Patent Citations over Time and across Institutional and Geographic Boundaries", *Proceedings of the National Academy of Sciences*, 93, pp. 12671 – 12677.

Harhoff, D., P. Régibeau and K. Rockett, 2001, "Some Simple Economics of GM Food", *Economic Policy*, October, pp. 265 – 299.

Henderson, R., A.B. Jaffe and M. Trajtenberg, 1998, "Universities as a Source of Commercial Technology: A Detailed Analysis of University Patenting", *Review of Economics and Statistics*, 80, pp. 119 – 127.

Lanjouw and M. Schankerman, , *NBER Working Papers*

Miller, R, and M.H. Davis, 1983, "*Intellectual Property: Patents, Trademarks and Copyright*", West Publishing Co., St Paul, Minnesota, pp. 428.

Régibeau, P. and K. Rockett, 2003, "Are Important Patents Approved More Slowly and Should They Be?", Working Paper, University of Essex.

Trajtenberg, M., R. Henderson and A.B. Jaffe, 1997, "University versus Corporate Patents: A Window on the Basicness of Invention", *Economics of Innovation and New Technology*, 5, pp. 9 – 50.